

## Article

# Semantic Positioning Model Incorporating BERT/RobERTa and Fuzzy Theory Achieves More Nuanced Japanese Adverb Clustering

Eric Odle <sup>1,\*</sup> , Yun-Ju Hsueh <sup>2</sup> and Pei-Chun Lin <sup>3</sup> 

<sup>1</sup> Department of Natural History Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

<sup>2</sup> Department of Foreign Languages and Applied Linguistics, Yuan Ze University, Taoyuan City 320315, Taiwan; connie@saturn.yzu.edu.tw

<sup>3</sup> Department of Information Engineering and Computer Science, Feng Chia University, Taichung City 407102, Taiwan; peiclin@mail.fcu.edu.tw

\* Correspondence: ericmichael.odle.q5@elms.hokudai.ac.jp

**Abstract:** Japanese adverbs are difficult to classify, with little progress made since the 1930s. Now in the age of large language models, linguists need a framework for lexical grouping that incorporates quantitative, evidence-based relationships rather than purely theoretical categorization. We herein address this need for the case of Japanese adverbs by developing a semantic positioning approach that incorporates large language model embeddings with fuzzy set theory to achieve empirical Japanese adverb groupings. To perform semantic positioning, we (i) obtained multi-dimensional embeddings for a list of Japanese adverbs using a BERT or RoBERTa model pre-trained on Japanese text, (ii) reduced the dimensionality of each embedding by principle component analysis (PCA), (iii) mapped the relative position of each adverb in a 3D plot using *K*-means clustering with an initial cluster count of  $n = 3$ , (iv) performed silhouette analysis to determine the optimal cluster count, (v) performed PCA and *K*-means clustering on the adverb embeddings again to generate 2D semantic position plots, then finally (vi) generated a centroid distance matrix. Fuzzy set theory informs our workflow at the embedding step, where the meanings of words are treated as quantifiable vague data. Our results suggest that Japanese adverbs optimally cluster into  $n = 4$  rather than  $n = 3$  groups following silhouette analysis. We also observe a lack of consistency between adverb semantic positions and conventional classification. Ultimately, 3D/2D semantic position plots and centroid distance matrices were simple to generate and did not require special hardware. Our novel approach offers advantages over conventional adverb classification, including an intuitive visualization of semantic relationships in the form of semantic position plots, as well as a quantitative clustering “fingerprint” for Japanese adverbs that express vague language data as a centroid distance matrix.

**Keywords:** bidirectional encoder representations from transformers; robustly optimized BERT approach; natural language processing; deep learning; large language models; fuzzy logic; silhouette analysis; principle component analysis



**Citation:** Odle, E.; Hsueh, Y.-J.; Lin, P.-C. Semantic Positioning Model Incorporating BERT/RobERTa and Fuzzy Theory Achieves More Nuanced Japanese Adverb Clustering. *Electronics* **2023**, *12*, 4185. <https://doi.org/10.3390/electronics12194185>

Academic Editor: George A. Tsihrintzis

Received: 1 September 2023

Revised: 26 September 2023

Accepted: 5 October 2023

Published: 9 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In this study, we focus on how an innovative BERT workflow can bring clarity to a notoriously confusing topic for Japanese linguists: adverbs. Recently, the field of Natural Language Processing (NLP) has seen rapid advancements, leading to powerful linguistic analysis tools [1,2]. Language modeling by machine learning [3,4], specifically using deep neural networks [5,6], has demonstrated superior performance on a multitude of linguistic tasks such as sentence prediction [7], question answering [8,9], and machine translation [10,11]. While these tasks differ, they share a basic mechanism

that involves converting text into a numerical representation (tokenization), sending that numerical representation through a series of mathematical modifications (forward propagation), then outputting a changed mathematical representation at the end of the model (embedding).

Large language models are presented with a massive amount of textual data during their training phases. While training, language models learn to make predictions, and in doing so they develop an understanding of grammar and vocabulary. Word embeddings, manifested as multi-dimensional vectors, therefore reflect a precise semantic position in high-dimensional space as learned by a given model on a given corpus of training text. Despite the free availability of this powerful technology, there is a disconnect between word embedding theory and practical application by linguists. This disconnect can be attributed to two factors: a high technical barrier required to manipulate language models in a programming environment, and a lack of established use cases. The present study aims to address both of these factors by offering a novel application of word embeddings to Japanese adverbs along with beginner-friendly Jupyter Notebooks for users to implement independently.

### *1.1. Approaches to Word Embedding*

Word embeddings encapsulate incredible knowledge using relatively few numbers, making them ideal for relational tasks such as vocabulary clustering. Methods of word embedding generation vary, but they all aim to satisfy three criteria: unique representation, low dimensionality, and semantic proximity reflected by position in  $n$ -dimensional space [12]. Such approaches to word embedding include Word2Vec [13] and GloVe [14], which are statistical approaches that represent words as high-dimensional vectors. Alternatively, word embeddings may be obtained from pre-trained language models such as recurrent neural networks or transformer-based BERT models. These approaches have been applied to sentiment analysis tasks, revealing Word2Vec, GloVe [15], and convolutional neural networks [16] to be effective embedding approaches in determining positive versus negative online discourse. Moreover, BERT and RoBERTa were previously used to visualize three-dimensional word embeddings in a recent study on fake news detection [17]. We herein build upon previous work by also visualizing language model word embeddings for the novel application of Japanese adverb clustering.

### *1.2. Conventional Japanese Adverb Classification*

Japanese adverbs are notoriously difficult to classify. Proposed classification schemes include pragmatic, evaluative, modal, and domain adverb classes [18], mood, tense, aspect, voice, and object-referential adverb classes [19], classification by に (ni), も (mo), or にも (ni-mo) particle conjugation [20], and a 41 computational class scheme [21]. Despite the wealth of alternative categorization approaches, Japanese adverbs are most often categorized using Yamada's [22] Degree/Declarative/Status scheme.

Beyond the simple categorization shown in Table 1, there have been few attempts to computationally analyze Japanese adverbs. One study [23] made use of IPADIC [24] and ChaSen [25] to extract a glossary of adverbs with accompanying connotation details. More recently, researchers considered the range of strength and polarity among Japanese Degree adverbs in the context of Japanese-to-Korean machine translation [26], highlighting a real-world need in the linguistics services industry for better computational models. Still, few studies focus on either the semantic position or clustering of Japanese adverbs from a computational perspective.

### *1.3. BERT-Based Language Models*

BERT is a 110-million parameter Large Language Model (LLM) published in 2019 [27]. The architecture within BERT is distinct from other LLMs in that it does not employ recurrence. Rather, BERT is composed of transformers [28], which are in turn built upon

multi-head attention layers made of a parallel series of scaled dot-product attention layers. A scaled dot-product attention layer may be described mathematically as

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

where  $\mathbf{Q}$ ,  $\mathbf{K}$ , and  $\mathbf{V}$  are matrices containing matching query, key, and value data, respectively. The superscript T is a linear algebra notation indicating that the preceding matrix is transposed. The element  $d_k$  represents the dimensions of the  $\mathbf{K}$  matrix, and division by the square of this value is used as a scaling factor in the softmax calculation. Scaled dot-product attention layers are combined in parallel to produce multi-head attention layers. Mathematically, the increased layering is expressed via the equation

$$MultiHead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(head_1, \dots, head_n)\mathbf{W}^O \quad (2)$$

for which

$$head_i = Attention(\mathbf{QW}_i^Q, \mathbf{KW}_i^K, \mathbf{VW}_i^V) \quad (3)$$

where the parameter matrix  $\mathbf{W}^O$  belongs to the set of all real numbers contained within the cross product of  $\mathbf{V}$  and the model, times the number of layers, while  $\mathbf{QW}_i^Q$ ,  $\mathbf{KW}_i^K$ , and  $\mathbf{VW}_i^V$  are similarly-defined parameter matrices for input data matrices  $\mathbf{Q}$ ,  $\mathbf{K}$ , and  $\mathbf{V}$ , respectively.

**Table 1.** Adverb category examples. Examples of the three Japanese adverb types according to the Yamada taxonomy from 1936. Declarative adverbs tend to express emphasis and uncertainty, Degree adverbs tend to express abstract and physical quantity, and Status adverbs tend to express auxiliary information pertaining to a subject such as its good/bad quality, state of activity, etc.

Type	Examples
Degree	少し ( <i>sukoshi</i> ; a little), 沢山 ( <i>takusan</i> ; much), とても ( <i>totemo</i> ; very), だいぶ ( <i>daibu</i> ; considerably)
Declarative	必ず ( <i>kanarazu</i> ; must/certainly), お陰で ( <i>okagede</i> ; thanks to), 実は ( <i>jitsuha</i> ; in fact), 多分 ( <i>tabun</i> ; probably so)
Status	よく ( <i>yoku</i> ; well), まだ ( <i>mada</i> ; not yet), 自分で ( <i>jibunde</i> ; by yourself), ずっと ( <i>zutto</i> ; constantly)

This transformer-based architecture advanced the capabilities of LLMs, and Japanese-trained BERT models continue to demonstrate utility in multiple fields such as medicine [29–32], literature [33,34], law [35], automation [36,37], and second language education [38,39]. Together, recent work suggests that BERT remains relevant, powerful, adaptable, and applicable in yet-unstudied ways. With the success of BERT, researchers began modifying the architecture and training parameters for optimizing various tasks. One such BERT-inspired model is RoBERTa (Robustly optimized BERT approach), which maintains the base BERT architecture but changes the training method to include dynamic rather than static masking, full-sentence training text without using next-sentence prediction loss, larger mini-batches, and a higher capacity text encoding scheme [40]. In this study, both base and large iterations of BERT and RoBERTa were considered during adverb analysis. Neither fine-tuning nor architectural modifications were performed in this study. Fine-tuning was avoided because there is no established ground truth to guide model training. Changes to the pre-trained architectures were avoided out of conflict with our stated goal of making the proposed process as user-friendly as possible.

#### 1.4. Fuzzy Logic in Natural Language Processing

Fuzzy logic, first proposed in 1965 [41], is a method of handling uncertainty by allowing the truth values to range from 0 to 1. This approach differs from Boolean logic,

which only allows for binary truth values of 0 (False) or 1 (True). Vagueness in the context of fuzzy logic refers to the ambiguous boundaries between categories and is a fundamental principle in fuzzy logic. Moreover, vagueness deserves attention because it reflects the complexity and nuance of real-world systems such as human language. While some scholars have argued that fuzzy logic is not well suited for linguistic semantic analysis [42,43], these conclusions were drawn without considering LLMs. More recently, researchers have revisited fuzzy logic as it applies to natural language. For example, fuzzy natural language [44] is a subfield that aims to model linguistic semantics using a fuzzy logic approach [45].

Within the field of Japanese linguistics, applications of fuzzy logic include studies on Japanese language pedagogy [46] and machine translation [47]. In line with our focus, some researchers echo the position that Japanese words may be best described through the lens of fuzzy logic rather than rigid a priori classification schemes [48]. However, to our knowledge, no published work has proposed a method for quantifying Japanese adverbs that takes into account the vagueness of human language. Fortunately, LLMs trained on massive text corpora are now freely available, providing researchers with high-dimensional word representations.

In this study, we define a semantic position model that uses the coordinates of a 768- or 1024-dimension Japanese adverb embedding following dimensional reduction via PCA. Using this simple and powerful 2-step process of embedding followed by dimensional reduction, we were able to obtain quantitative representations of learned meaning for Japanese adverbs. This semantic position model enabled us to perform two key tasks: (1) evaluate the conventional Japanese adverb classification scheme against LLMs, and (2) propose a novel quantification approach to working with Japanese adverbs that researchers and language teachers in the age of computational linguistics will find useful. Finally, we conclude that our semantic position model, informed by fuzzy logic, better reflects the vague data that is human language.

## 2. Methodology

In this study, we present a novel semantic position model for Japanese adverbs inspired by fuzzy logic and driven by LLM word embedding. In the most general sense, our model takes as input a list of Japanese adverbs in plain text form, then generates two products as the output: (1) a 2D semantic position plot with number of clusters optimized by Silhouette Analysis, and (2) a centroid distance matrix of the position plot. A pseudo-code outline of the proposed workflow is provided below in Algorithm 1.

---

### Algorithm 1 Description of the Workflow

---

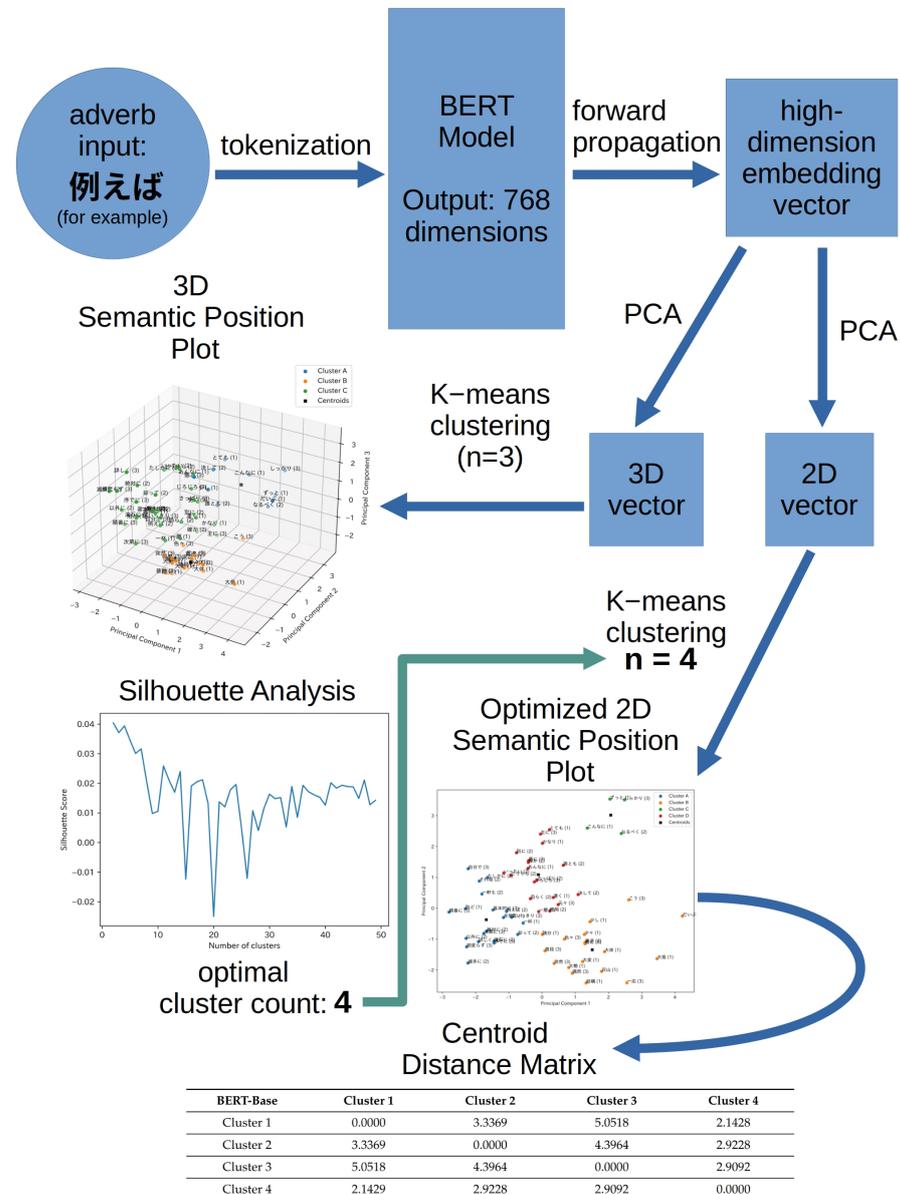
**Input:** List of Japanese adverbs in plain text  
**Output:** 2D semantic position plot and centroid distance matrix

- 1: **procedure**
- 2:   Generate word embeddings by LLM
- 3:   Reduce dimensionality with PCA
- 4:   Initialize  $K$ -means Clustering with 3 clusters
- 5:   Plot adverb positions in 3D
- 6:   Determine optimal cluster count by Silhouette Analysis
- 7:   Re-cluster with the optimal cluster count
- 8:   Plot adverb positions in 2D
- 9:   Generate centroid distance matrix
- 10: **end procedure**

---

In further detail, this semantic positioning model involves the following steps. First, we obtain multi-dimensional embeddings by passing a list of Japanese adverbs through a LLM such as BERT or RoBERTa. Second, we reduce the dimensionality of each embedding by principle component analysis (PCA), then plot the positions of each adverb in 3D following  $K$ -means clustering with an initial cluster count of  $n = 3$ . Third, we perform

silhouette analysis to determine the optimal cluster count, followed by another round of PCA and K-means clustering on the original embedding vector to generate a 2D plot. Finally, we generate a centroid distance matrix for each 2D plot. The analysis workflow is shown visually as a flowchart in Figure 1.



**Figure 1.** Workflow illustrating the proposed semantic positioning model for Japanese adverbs incorporating LLM embedding, principle component analysis, and K-means clustering. The Japanese adverb 例え (meaning “for example”) is provided as an example input.

2.1. Adverb Selection

Following Yamada’s Degree/Declarative/Status classification scheme [22], an initial list of 350 common adverbs was collected and categorized. Categories were unevenly represented, so each category was randomly assigned a more manageable size of 20 adverbs. Written Japanese incorporates Chinese characters (kanji), resulting in multiple variants (akin to alternative spellings in English) for many words. Given that BERT models are trained on written text, we aimed to select the adverb variants most likely to appear in writing. The resulting adverb list is shown in Table 2.

**Table 2.** List of adverbs used for analysis. This table shows 20 Japanese adverbs from the categories Degree, Declarative, and Status, following Yamada’s [22] three-category classification system. Adverbs shown in this table were used to evaluate the semantic position model.

Degree	Declarative	Status
一層	以外に	一応
非常に	一軒も	序でに
凄く	いったい	相変らず
結構	恐らく	思いっきり
だいぶ	却って	主に
あんなに	決して	普段
少々	流石に	基本的に
ずっと	さっぱり	態と
大変	確か	偶然
一杯	それ程	自分で
かなり	誰とも	詳しく
こんなに	どうやら	こう
大抵	なるべく	更に
殆ど	真逆	突然
大勢	減多に	しっかり
大体	別に	元々
随分	例えば	次第に
沢山	結局	色々
少し	たしかに	順番に
とても	絶対に	じろじろ

## 2.2. Model Selection

The training process for a complex model such as BERT can be computationally prohibitive, requiring a GPU and extended run-time resources. Fortunately, several general-purpose BERT models pre-trained on Japanese text are freely available. Four models were selected for this study: the BERT-base-Japanese and BERT-large-Japanese models from Tohoku University [49], and the RoBERTa-base-Japanese and RoBERTa-large-Japanese models from Waseda University [50]. These BERT models were trained on massive amounts of internet text, providing broad linguistic exposure. Such pre-trained knowledge can be invaluable for tasks where understanding context and semantics is critical, as in word embedding. Moreover, BERT/RoBERTa were selected over generative pre-trained transformer (GPT) models for their encoder-decoder architecture, which allows for contextual learning in both left-to-right as well as right-to-left directions. BERT-base-Japanese was trained on approximately 30 million Japanese sentences pulled from Wikipedia. The training lasted for 5 days using TPU acceleration, processing 512 tokens per pass with a batch size of 256 and a total of 1 million training epochs [49]. The RoBERTa models were trained on a Japanese Wikipedia dump in addition to the Japanese portion of the CC-100 corpus [51] over seven days using eight NVIDIA A100 GPUs [50].

## 2.3. Semantic Positioning: 3D Plotting and Comparison

Our first data manipulation step aimed to test the validity of the conventional Degree/Declarative/Status scheme against massive language models. This experiment follows the reasoning that if a three-category classification truly reflects semantic relationships among Japanese adverbs, then those relationships should be quantitatively apparent

through LLMs fed more text than is feasible for human reading. For each of the four models examined (BERT-base, BERT-large, RoBERTa-base, and RoBERTa-large), adverbs were tokenized using the model's native tokenizer, then passed through the model to generate a multi-dimensional embedding for each adverb. Specific embedding dimensions depend on the model used: BERT-base/RoBERTa-base has an output layer of 768 dimensions, while BERT-large/RoBERTa-large has an output layer of 1024 dimensions. These embedding values were then normalized to between 0 and 1, compressed to 3 principle components using PCA, and then clustered using  $K$ -means clustering. Finally, these processed adverb representations were plotted in 3D [52] to visualize their relative positions in compressed semantic space.

PCA and  $K$ -means clustering were performed using scikit-learn [53], a popular machine learning library for Python. Principle component analysis (PCA) [54] is a method of reducing the dimensionality of a dataset without losing essential differences (variance) between each point. Dimensionality is reduced by searching for perpendicular (orthogonal) axes called principle components that preserve the highest variance among the dataset. PCA involves five main steps: centering, covariance matrix calculation, eigenvalue/eigenvector calculation on the covariance matrix, eigenvector sorting, component selection, and low-dimensional projection. The covariance matrix  $\mathbf{C}$  for a dataset is calculated by multiplying the data feature matrix  $\mathbf{X}$  ( $m$  data points,  $n$  dimensions, and shape  $m \times n$ ) with its transposed clone  $\mathbf{X}^T$  according to the below formula.

$$\mathbf{C} = \frac{1}{m-1} \mathbf{X} * \mathbf{X}^T \quad (4)$$

During the subsequent eigenvalue/eigenvector calculation step, eigenvalues ( $\lambda$ ) and eigenvectors ( $v$ ) are calculated for each dimension of the dataset as solutions to the equation

$$\mathbf{C}v = \lambda v \quad (5)$$

Eigenvalues reflect the amount of variance in each eigenvector, while the eigenvectors themselves reflect an axis oriented in the original high-dimensional dataset space. In this study, we decomposed high-dimensional Japanese adverb embeddings from over 700 dimensions down to either 2 or 3 dimensions. With respect to PCA, this means we used the two/three eigenvectors with the largest eigenvalues for our semantic position model.

Meanwhile,  $K$ -means clustering [55] is a process that aims to group points in a dataset into a set number of clusters by minimizing the variance within each cluster. In general,  $K$ -means clustering follows the steps of setting a number of clusters, assigning initial cluster centroid positions, calculating distance (Euclidean) from each centroid for each point in the dataset, assigning each point to its nearest cluster centroid, repositioning each centroid to the mean position of all the points in its cluster, and checking for convergence to either exit or repeat the process. During cluster assignment, the process may be expressed mathematically in terms of points  $x_i$  and centroid  $c_j$ . The  $K$ -means algorithm searches for an index value  $j$  that minimizes the Euclidean distance to cluster  $j$  from 1 to  $K$  clusters. Finally,  $x_i$  is assigned to cluster  $j$  in line with the below formula.

$$\text{Membership to cluster } j = \operatorname{argmin} \|x_i - c_j\| \quad (6)$$

This is followed by the centroid update step, where the position of centroid  $c_j$  for each cluster  $j$  is relocated to the mean position of every point  $i$  in the cluster of  $N_j$  total points according to

$$\text{Updated } c_j = \frac{1}{N_j} \sum_i x_i \quad (7)$$

After  $K$ -means clustering, we are left with Japanese adverb embeddings optimally grouped into a pre-defined number of clusters.

#### 2.4. Silhouette Analysis

A central question asked in this study is “What is the optimal number of categorical clusters for Japanese adverbs?” To address this question, silhouette analysis [56] was performed on the principle components of our embedded adverb vectors. Silhouette scores from 1 to 50 clusters were considered during analysis, although only scores for clusters of  $n = 3$  or larger were used as possible valid results. Silhouette analysis is a technique used to evaluate the quality of clustering results from unsupervised machine learning algorithms such as  $K$ -means. Specifically, silhouette analysis helps determine how well-separated clusters are and whether data points within each cluster are more similar to one another than they are to data points in other clusters.

As originally proposed [56], the silhouette score  $s(i)$  for a single point is calculated by

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

for which  $a(i)$  represents the mean Euclidean distance of point  $i$  to other points shared by the same cluster, and  $b(i)$  represents the minimum mean Euclidean distance found between point  $i$  and another point belonging to a different cluster. From this, the silhouette score for a set of points can be defined as the mean of all silhouette scores for those points using the expression

$$\text{Silhouette Score} = \frac{1}{N} \sum_{i=1}^N s(i) \quad (9)$$

where  $N$  is the number of points in a given set. In our study,  $s(i)$  was calculated for each three-dimensional Japanese adverb embedding following PCA. Then, the silhouette score was calculated for the entire  $N = 60$  set of adverbs.

#### 2.5. Semantic Positioning: 2D Plotting and Distance Matrix Construction

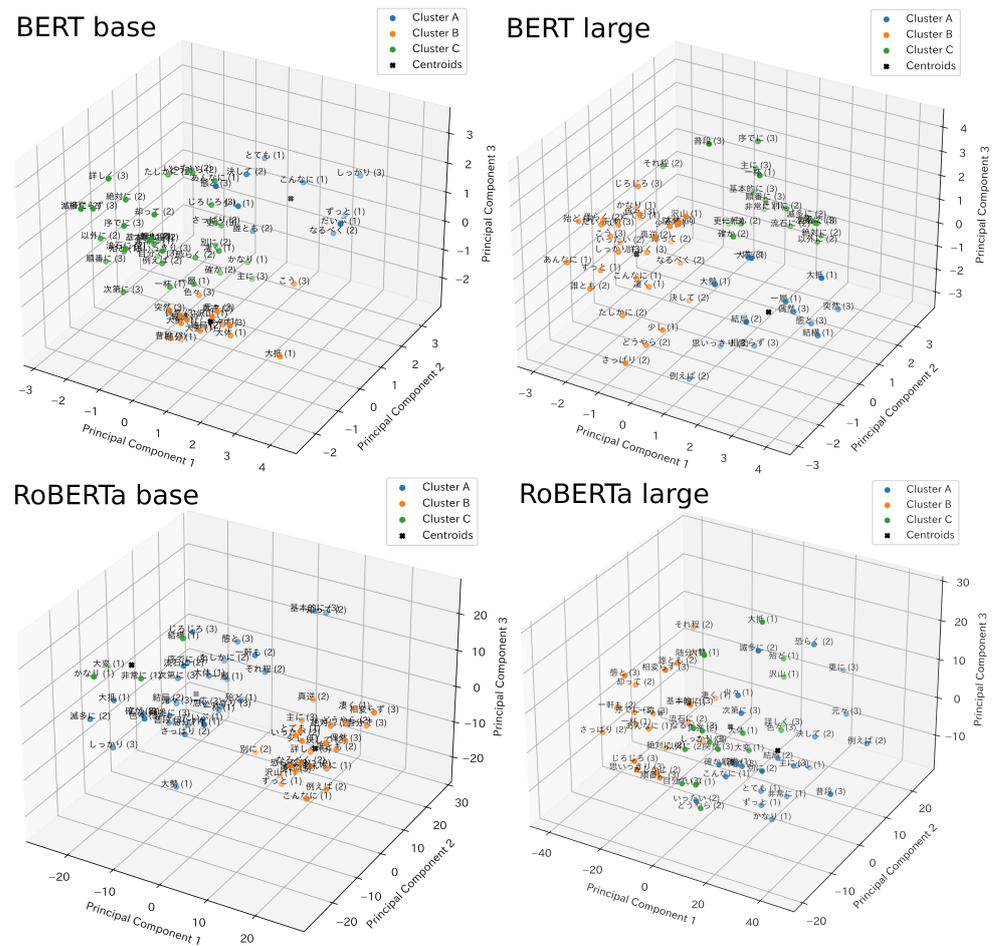
Following silhouette analysis, raw adverb embeddings were re-compressed down to two dimensions via PCA for 2D plotting. Next, the optimal number of clusters as determined by silhouette analysis was used to repeat  $K$ -means clustering. Resulting cluster centroids were plotted along with adverb embeddings to visualize the revised categorization scheme, then a matrix was constructed to express the distances from each centroid to the other. This distance matrix represents the final product of our semantic position model: a clustering “fingerprint” for Japanese adverbs.

### 3. Experiment Results

During this study, we saw three major results. First, the conventional categorization system (Degree/Declarative/Status) for Japanese adverbs did not agree with the semantic positions from any of the four tested BERT/RobERTa models. Second, while semantic positions varied by language model,  $n = 4$  clusters fit the data better than  $n = 3$  in all four BERT/RobERTa models. Third, we were able to generate distance matrix “fingerprints” for our target set of Japanese adverbs specific to individual BERT/RobERTa models. These representations of vague human data draw inspiration from fuzzy theory, providing researchers with a novel, natural language-based framework for Japanese adverb classification.

#### 3.1. Conventional Categories Do Not Reflect Semantic Position

Adverb embedding using four transformer-based LLMs trained on Japanese text—BERT-base, BERT-large, RoBERTa-base, and RoBERTa-large—revealed no apparent overlap between the Yamada classification scheme and 3D principal component analysis followed by  $K$ -means clustering. These results are visualized in Figure 2 below.



**Figure 2.** Adverb representations in 3D semantic space following BERT/RoBERTa embedding, principle component analysis (three principal components), and *K*-means clustering ( $n = 3$ ). Points are labeled with corresponding adverb, written in Japanese. Conventional categories Degree, Declarative, and Status are represented in parentheses beside adverbs as 1, 2, and 3, respectively. *K*-means clustering results are expressed by the colors blue, yellow, and green. Cluster centroids are shown as bold marks.

Adverb semantic positions are shown in Figure 2 accompanied by Japanese text, conventional category labels (Degree = 1, Declarative = 2, Status = 3; in parentheses), and *K*-means cluster associations (blue, yellow, and green; cluster centroids indicated as bold marks). There was minimal overlap between conventional classes and semantic embedding by the four Japanese language models tested. Plot structure, including both cluster centroid and adverb positions, varied widely depending on language model. Moreover, cluster tightness varied greatly, with large models tending to produce more disperse semantic position clustering.

As seen in Table 3, the scores were generally low across four common classification metrics for all four LLMs considered. The RoBERTa models tended to perform better than the BERT models, and base size models outperformed large size models. None of the four models tested were able to achieve a classification accuracy of 0.5 or higher, suggesting poor overlap between Yamada’s [22] a priori classification scheme and empirically-derived semantic positions.

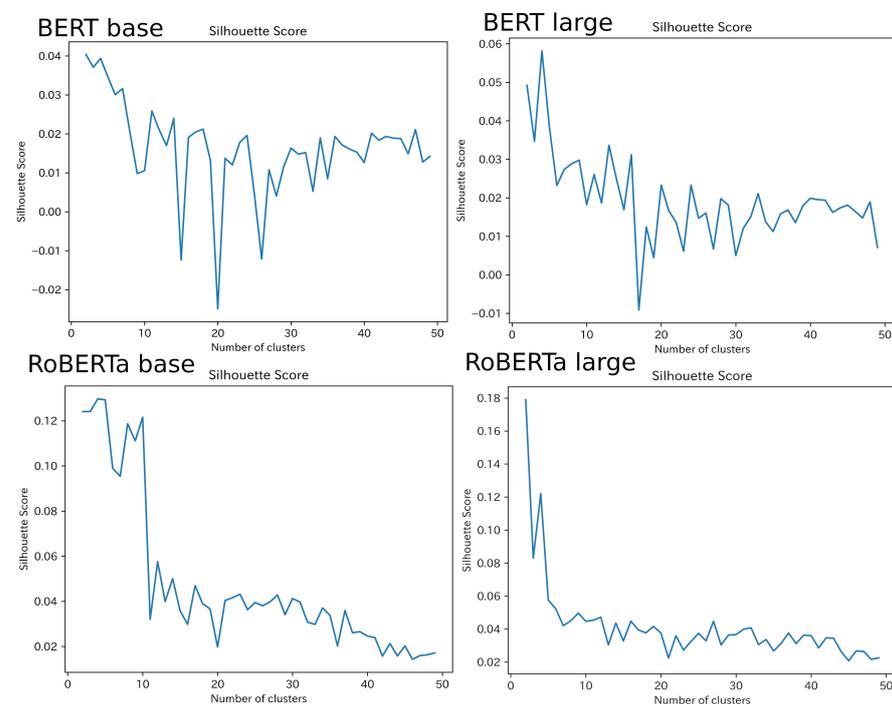
**Table 3.** Classification scores from four Japanese Language models. This table provides scores for four common classification task metrics: Recall, Precision, F1-Score, and Accuracy. Scores were calculated based on Yamada’s [22] conventional degree/declarative/status classifications scheme for Japanese adverbs.

Model	Recall	Precision	F1-Score	Accuracy
BERT base	0.367	0.368	0.349	0.367
BERT large	0.317	0.345	0.323	0.317
RoBERTa base	0.483	0.632	0.458	0.483
RoBERTa large	0.400	0.403	0.399	0.400

### 3.2. Four Adverb Categories are Better than Three

Silhouette analysis was performed on cluster counts ranging from 1 to 50 to better grasp clustering trends in the data. For all models, silhouette scores tended to decrease with the increasing cluster count. With the RoBERTa models, a drastic drop-off in silhouette score was seen within the first 10 cluster sizes, whereas the BERT models decreased less dramatically. Moreover, while all four models displayed a degree of saw-like peaks and valleys, RoBERTa models produced much smoother silhouette score curves than the BERT models. The BERT-base in particular experienced an intermediate period of sharp rises and falls, even into negative values, before completing the analysis cycle near the top of the score range.

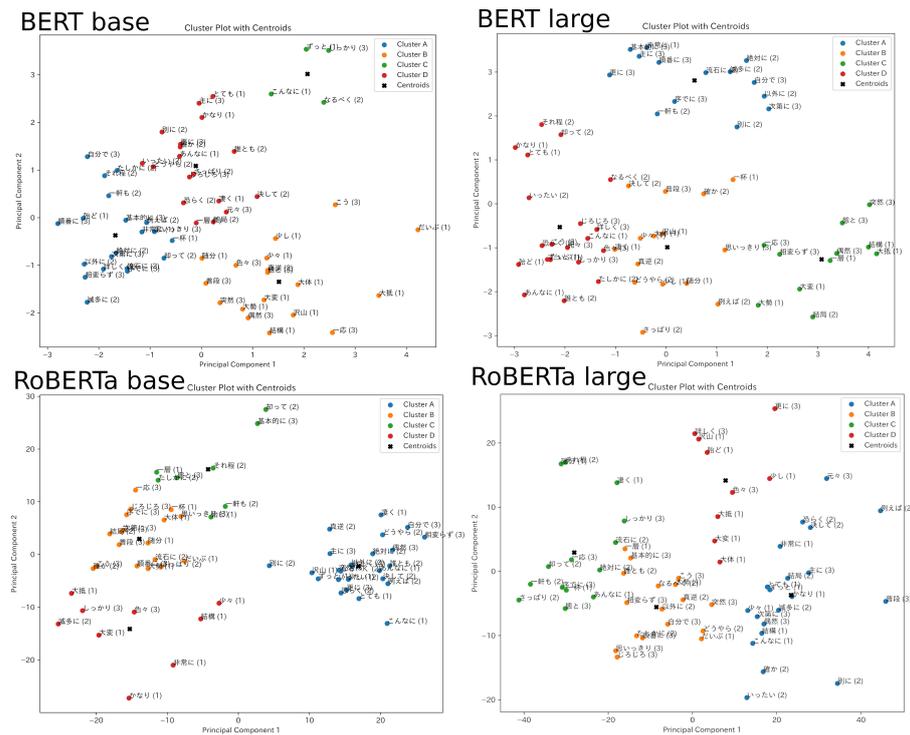
Ultimately, silhouette analysis found  $n = 4$  clusters to better fit our adverb embeddings than  $n = 3$  clusters for all four models considered. Given the results shown in Figure 3, a clustering scheme using  $n = 4$  was employed for subsequent analysis.



**Figure 3.** Line plots of silhouette score vs. cluster number for four Japanese language models: BERT-base, BERT-large, RoBERTa-base, and RoBERTa-large. In all four models,  $n = 4$  clusters produced the highest silhouette score (excluding  $n = 1$  and  $n = 2$ ).

### 3.3. Novel Framework for Classifying Japanese Adverbs

As with the 3D plots generated prior to silhouette analysis, conventional categorization did not overlap well with either BERT or RoBERTa embeddings. Similarly, adverb positions among the four models, even after re-clustering into the optimal  $n = 4$  clusters, still showed a high degree of heterogeneity between models. This is visually apparent in Figure 4 below.



**Figure 4.** Adverb representations in 2D semantic space following BERT/RobERTa embedding, principle component analysis (two principle components), and *K*-means clustering (*n* = 4). Points are labeled with corresponding adverb, written in Japanese. Conventional categories Degree, Declarative, and Status are represented in parentheses beside adverb markers as 1, 2, and 3, respectively. *K*-means clustering results are expressed by the colors blue, yellow, and green. Cluster centroids are shown as bold marks.

In this study, we also introduced the novel approach of generating a centroid distance matrix for Japanese adverbs following semantic positioning. Example distance matrices are shown in Table 4, and present the relative Euclidean distances between each cluster centroid (*n* = 4) following model embedding and dimensional reduction by PCA (principle components = 2).

**Table 4.** Adverb centroid distance matrices from four Japanese language models. This table presents relative Euclidean distances between cluster centroids (*n* = 4) for each of the four Japanese language models considered (BERT-base, BERT-large, RoBERTa-base, and RoBERTa-large) following model embedding, PCA, and *K*-means clustering.

BERT-base	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0.0000	3.3369	5.0518	2.1428
Cluster 2	3.3369	0.0000	4.3964	2.9228
Cluster 3	5.0518	4.3964	0.0000	2.9092
Cluster 4	2.1429	2.9228	2.9092	0.0000
BERT-large	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0.0000	3.8363	4.7875	4.2708
Cluster 2	3.8363	0.0000	3.0638	2.1723
Cluster 3	4.7875	3.0638	0.0000	5.2266
Cluster 4	4.2708	2.1723	5.2265	0.0000

**Table 4.** *Cont.*

<b>RoBERTa-base</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>
Cluster 1	0.0000	31.2850	28.0213	34.3317
Cluster 2	31.2850	0.0000	16.3790	17.1308
Cluster 3	28.02123	16.3790	0.0000	32.2291
Cluster 4	34.3317	17.1308	32.2291	0.0000
<b>RoBERTa-large</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>
Cluster 1	0.0000	32.0137	51.8964	23.6767
Cluster 2	32.0137	0.0000	21.2812	25.6508
Cluster 3	51.8964	21.2812	0.0000	37.6306
Cluster 4	23.6767	25.6508	37.6306	0.0000

The four matrices in Table 4 were obtained using our proposed semantic position model for Japanese adverbs. First, adverbs were embedded using one of four LLMs. Next, high-dimensional embeddings were projected onto a 2D plane by PCA, and then  $K$ -means clustering was used ( $n = 4$ ) to group the projected embeddings. PCA allowed for the reduction of the high-dimensional textual data into a lower-dimensional space, preserving the most important variance between data points while minimizing noise. Cluster centroids are positioned from one another at varying proportions depending on the particular language model used (BERT-base, BERT-large, RoBERTa-base, or RoBERTa-large).

#### 4. Discussion

The primary aim of this study was to evaluate a conventional Japanese adverb classification system using LLMs. Informed by the fuzzy set theory concept of vague data, we arrived at a novel model for Japanese adverbs referred to as semantic positioning. First, our  $n = 3$  cluster test showed minimal overlap between conventional categorical delineation and embedding-based clustering among Japanese adverbs when tested on four Japanese language models (BERT-base, BERT-large, RoBERTa-base, and RoBERTa-large). Quantitatively, RoBERTa models reproduced the Yamada [22] classification scheme with better accuracy than BERT models, although none of the models achieved an accuracy of 0.5 or greater. The slight advantage seen with RoBERTa over BERT is attributed to the larger training text dataset. Model size did not provide a clear classification accuracy advantage, with base size models outperforming their respective large size models (Table 3). Together, these findings suggest that the a priori Degree/Declarative/Status classification scheme is fundamentally incompatible with empirical, context-rich LLM word embeddings. Intuitively, this result is unsurprising. Language in everyday use does not follow rigid rules. Human thoughts are difficult to quantify—as are culture, intelligence, and other factors. For this reason, the fuzzy logic notion of vague data was well suited for our subject matter: Japanese adverbs.

Another contribution made by our study was in demonstrating the utility of LLMs such as BERT and RoBERTa as engines for semantic embedding. Semantic distance, generally defined as the differences between words using a pre-determined metric [57], was used in our study to quantify the relationships between Japanese adverbs. Semantic distance is a result of passing text through a language model via forward propagation, applying the weights and biases at each transformer layer to the tokenized text. Conceptually, if two words are passed through BERT, then the distance between the resulting vector embeddings can be calculated by simple Euclidean means [58]. Ideally, those embeddings should be closer to each other the more similar the original words are in meaning—hence, semantic position.

One more interesting finding from this study was the striking inconsistency in adverb positions and cluster behavior following plot generation (3D in Figure 2 and 2D in Figure 4). This could be due to any of multiple variables including but not limited to model

size, training data size, word frequency within the training data, tokenization methods, random initialization, and hyperparameter differences. The four BERT/RobERTa models all varied in these respects, sufficiently explaining the differences seen in our results. As such, researchers and educators should mind the following point when implementing our semantic position model: keep the language model consistent. This inconsistency between models has been noted by other researchers as well. For example, a comparative study of 100 BERT models trained on the same dataset found consistent genre identification capabilities among models but wide variation in the ability to generalize text [59]. Overall, LLM complexity has outpaced our ability to fully describe their behavior, thus more work is needed in this area.

## 5. Conclusions and Future Work

In this study, we evaluated the Degree/Declarative/Status scheme of Japanese adverb classification using multi-dimensional embeddings obtained from LLMs. In doing so, we devised an alternative framework which we called a semantic positioning model. Our workflow involved (i) obtaining multi-dimensional embeddings for a pre-defined list of Japanese adverbs using a BERT/RobERTa model, (ii) reducing the dimensionality of each embedding by PCA, (iii) mapping relative positions for each adverb in a 3D plot using  $K$ -means clustering with an initial cluster count of  $n = 3$ , (iv) performing silhouette analysis to determine optimal cluster count, (v) performing PCA and  $K$ -means clustering again on the adverb embeddings to generate 2D semantic position plots, and (vi) generating centroid distance matrices. These final distance matrices serve as unique identifying “fingerprints” for Japanese adverbs.

Our novel process revealed three key findings: (1) Japanese adverbs optimally clustered into  $n = 4$  rather than  $n = 3$  groups following silhouette analysis, (2) there was little consistency between semantic positions and conventional classifications, and (3) plots and centroid distance matrices were simple to generate without the use of special hardware.

From these results, we arrived at the following conclusions. First, our novel semantic positioning workflow offers a simple and powerful method for understanding Japanese adverbs as points in semantic space. This quantification of vague human data gives linguists a more flexible way to work with language. Semantic positions varied widely from model to model, and our method did not attempt to re-categorize adverbs based on cluster membership alone. As such, further syntactic analysis will likely be required to reconstruct a rigid four-class categorization system for Japanese adverbs in the future. Second, we provided quantitative experimental evidence suggesting the conventional Degree/Declarative/Status scheme does not hold true when compared against LLMs such as BERT or RobERTa, which are trained on massive amounts of real-world text. This finding will hopefully inspire other researchers to abandon the conventional classification scheme for Japanese adverbs, or at least preface the use of the scheme by emphasizing its theoretical nature.

Furthermore, our investigation highlighted an important consideration when working with LLMs: inter-model variability. The inconsistency in semantic positions from model to model was striking and echoed the observations made by other researchers in the field. In summary, we have herein proposed a simple, powerful, and quantitative workflow for understanding Japanese adverbs as they manifest in natural language. It is our hope that linguists and educators consider integrating our workflow, provided as easy-to-follow Jupyter Notebooks, during research or Japanese language instruction.

**Author Contributions:** Conceptualization, E.O. and Y.-J.H.; methodology, E.O.; investigation, E.O. and Y.-J.H.; writing—original draft preparation, E.O. and P.-C.L.; writing—review and editing, P.-C.L.; visualization, E.O.; supervision, Y.-J.H.; funding acquisition, E.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Ministry of Education, R.O.C., under the grants of TEEP@AsiaPlus and the MOE Teaching Practice Research Program, 2023.

**Data Availability Statement:** The Python code used to generate our results are provided as Jupyter Notebooks for both BERT and RoBERTa models on the project’s GitHub repository: <https://github.com/ericodde/Semantic-Positioning-Model-for-Japanese-Adverbs> (accessed on 1 August 2023). Source text and English translations for Japanese adverbs are also provided in the repository.

**Acknowledgments:** We would like to extend our sincere gratitude to Kazuhiro Kogame, and Kevin C. Wakeman for allowing the finalization of this study to overlap with other ongoing projects. We would also like to thank the anonymous reviewers and the academic editor for their time, constructive suggestions, and insightful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

LLM	Large Language Model
NLP	Natural Language Processing
PCA	Principle Component Analysis
BERT	Bidirectional Encoder Representations from Transformers
RoBERTa	Robustly optimized BERT pre-training approach

## References

1. Hirschberg, J.; Manning, C.D. Advances in natural language processing. *Science* **2015**, *349*, 261–266. [CrossRef] [PubMed]
2. Omar, M.; Choi, S.; Nyang, D.; Mohaisen, D. Robust natural language processing: Recent advances, challenges, and future directions. *IEEE Access* **2022**, *10*, 86038–86056. [CrossRef]
3. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
4. Mahesh, B. Machine learning algorithms—A review. *Int. J. Sci. Res. IJSR* **2020**, *9*, 381–386.
5. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]
6. Arisoy, E.; Sainath, T.N.; Kingsbury, B.; Ramabhadran, B. Deep neural network language models. In Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT, Montréal, QC, Canada, 8 June 2012; pp. 20–28.
7. Bello, A.; Ng, S.C.; Leung, M.F. A BERT framework to sentiment analysis of tweets. *Sensors* **2023**, *23*, 506. [CrossRef]
8. Kierszbaum, S.; Lapasset, L. Applying distilled BERT for question answering on ASRS reports. In Proceedings of the 2020 New Trends in Civil Aviation (NTCA), Prague, Czech Republic, 23–24 November 2020; pp. 33–38.
9. Pandey, A.; Bhat, A. A Review on Textual Question Answering with Information Retrieval and Deep Learning Aspect. In Proceedings of the 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 17–19 May 2023; pp. 224–229.
10. Singh, S.P.; Kumar, A.; Darbari, H.; Singh, L.; Rastogi, A.; Jain, S. Machine translation using deep learning: An overview. In Proceedings of the 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur, India, 1–2 July 2017; pp. 162–167.
11. Imamura, K.; Sumita, E. Recycling a pre-trained BERT encoder for neural machine translation. In Proceedings of the 3rd Workshop on Neural Generation and Translation, Hong Kong, 4 November 2019; pp. 23–31.
12. Incitti, F.; Urli, F.; Snidaro, L. Beyond word embeddings: A survey. *Inf. Fusion* **2023**, *89*, 418–436.
13. Goldberg, Y.; Levy, O. word2vec Explained: Deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv* **2014**, arXiv:1402.3722.
14. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
15. Subba, B.; Kumari, S. A heterogeneous stacking ensemble based sentiment analysis framework using multiple word embeddings. *Comput. Intell.* **2022**, *38*, 530–559. [CrossRef]
16. Mangione, S.; Siino, M.; Garbo, G. Improving Irony and Stereotype Spreaders Detection using Data Augmentation and Convolutional Neural Network. In Proceedings of the CEUR Workshop Proceedings, Bologna, Italy, 5–8 September 2022; Volume 3180, pp. 2585–2593.
17. Siino, M.; Di Nuovo, E.; Tinnirello, I.; La Cascia, M. Fake news spreaders detection: Sometimes attention is not all you need. *Information* **2022**, *13*, 426. [CrossRef]
18. Nakau, M. *Ninchi Imiron no Genri (Principles of Cognitive Semantics)*; Taishukan: Tokyo, Japan, 1994.
19. Noda, H. Fukusi-no gojyun. *Nihongo Kyooiku* **1984**, *52*, 79–90.
20. Endo, Y. *Locality and Information Structure*; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2007.

21. Ogura, K.; Bond, F.; Ikehara, S. A method of ordering English adverbs. *J. Nat. Lang. Process.* **1997**, *4*, 17–39. [[CrossRef](#)] [[PubMed](#)]
22. Yamada, Y. *Nihon Bunpou Gaku Gairon (Survey of Japanese Grammar)*; Houbun Kan: Tokyo, Japan, 1936.
23. Kanamaru, T.; Murata, M.; Isahara, H. Construction of Adverb Dictionary that Relates to Speaker Attitudes and Evaluation of Its Effectiveness. In Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation, Wuhan, China, 1–3 November 2006; pp. 295–302.
24. Asahara, M.; Matsumoto, Y. ipadic version 2.7. 0 User’s Manual. Nara Institute of Science and Technology. 2003. Available online: <https://ja.osdn.net/projects/ipadic/docs/ipadic-2.7.0-manual-en.pdf/en/1/ipadic-2.7.0-manual-en.pdf.pdf> (accessed on 31 August 2023).
25. Matsumoto, Y.; Kitauchi, A.; Yamashita, T.; Hirano, Y.; Matsuda, H.; Takaoka, K.; Asahara, M. Japanese morphological analysis system ChaSen version 2.0 manual. In *NAIST Technical Report*; Nara Institute of Science and Technology: Ikoma, Japan, 1999.
26. Park, M. A Study on the Processing Pattern of Adverbs of Degree in Machine Translation—Focusing on the translation from Japanese to Korean—. *Japan Res.* **2023**, *59*, 69–90.
27. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017; Volume 30.
29. Kawazoe, Y.; Shibata, D.; Shinohara, E.; Aramaki, E.; Ohe, K. A clinical specific BERT developed with huge size of Japanese clinical narrative. *medRxiv* **2020**, *16*. Available online: <https://www.medrxiv.org/content/10.1101/2020.07.07.20148585v1> (accessed on 31 August 2023).
30. Araki, K.; Matsumoto, N.; Togo, K.; Yonemoto, N.; Ohki, E.; Xu, L.; Hasegawa, Y.; Satoh, D.; Takemoto, R.; Miyazaki, T. Developing artificial intelligence models for extracting oncologic outcomes from japanese electronic health records. *Adv. Ther.* **2023**, *40*, 934–950. [[CrossRef](#)] [[PubMed](#)]
31. Ohtsuka, T.; Kajiwara, T.; Tanikawa, C.; Shimizu, Y.; Nagahara, H.; Ninomiya, T. Automated Orthodontic Diagnosis from a Summary of Medical Findings. In Proceedings of the 5th Clinical Natural Language Processing Workshop, Toronto, ON, Canada, 14 July 2023; pp. 156–160.
32. Nishigaki, D.; Suzuki, Y.; Wataya, T.; Kita, K.; Yamagata, K.; Sato, J.; Kido, S.; Tomiyama, N. BERT-based Transfer Learning in Sentence-level Anatomic Classification of Free-Text Radiology Reports. *Radiol. Artif. Intell.* **2023**, *5*, e220097. [[CrossRef](#)] [[PubMed](#)]
33. Ueda, N.; Kawahara, D.; Kurohashi, S. BERT-based Cohesion Analysis of Japanese Texts. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 1323–1333.
34. Amano, M.; Tsumuraya, K.; Uehara, M.; Adachi, Y. *An Analysis of Representative Works of Japanese Literature Based on Emotions and Topics*; Springer: Cham, Switzerland, 2023; pp. 99–112.
35. Yamakoshi, T.; Komamizu, T.; Ogawa, Y.; Toyama, K. Japanese mistakable legal term correction using infrequency-aware BERT classifier. *Trans. Jpn. Soc. Artif. Intell.* **2020**, *35*, 4342–4351. [[CrossRef](#)]
36. Yawata, K.; Suzuki, T.; Kiryu, K.; Mohri, K. Performance Evaluation of Japanese BERT Model for Intent Classification Using a Chatbot. *Jpn. Soc. Artif. Intell.* **2021**, *35*.
37. Saito, Y.; Iimori, E.; Takamichi, S.; Tachibana, K.; Saruwatari, H. CALLS: Japanese Empathetic Dialogue Speech Corpus of Complaint Handling and Attentive Listening in Customer Center. *arXiv* **2023**, arXiv:2305.13713.
38. Ide, Y.; Mita, M.; Nohejl, A.; Ouchi, H.; Watanabe, T. Japanese Lexical Complexity for Non-Native Readers: A New Dataset. *arXiv* **2023**, arXiv:2306.17399.
39. Huy, P.T. Implementation of Automated Feedback System for Japanese Essays in Intermediate Education. Master’s Thesis, Japan Advanced Institute of Science and Technology, Nomi, Japan, 2023.
40. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
41. Zadeh, L.A. Fuzzy sets. *Inf. Control.* **1965**, *8*, 338–353. [[CrossRef](#)]
42. Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *J. Philos. Log.* **1973**, *2*, 458–508. [[CrossRef](#)]
43. Sauerland, U. Vagueness in language: The case against fuzzy logic revisited. In *Understanding Vagueness: Logical, Philosophical and Linguistic Perspectives*; College Publications: Rickmansworth, UK, 2011; pp. 185–198.
44. Novák, V. Fuzzy natural logic: Towards mathematical logic of human reasoning. In *Towards the Future of Fuzzy Logic*; Springer: Cham, Switzerland, 2015; pp. 137–165.
45. Novák, V. Fuzzy logic in natural language processing. In Proceedings of the 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy, 9–12 July 2017; pp. 1–6.
46. Yu, X.; Liu, X. Evaluation Method of Japanese Teaching Effect Based on Feature Offset Compensation. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 108. [[CrossRef](#)]
47. Chenhui, I.N.S.S.S.; Kurohashi, C.S. Filtering of a Web-Crawled Corpus to Achieve a Strong MT Model: A Case Study on the Japanese-Bulgarian Language Pair. In Proceedings of the 29th Annual Conference on Natural Language Processing (NLP2023), Okinawa, Japan, 13–17 March 2023.
48. Hoshi, H. Fuzzy Categories, Dynamic Labeling and Mixed Category Projections: The Case of Adjectival Nouns and Verbal Nouns. *秋田大学高等教育グローバルセンター紀要 [Akita University Global Center of Higher Education]* **2023**, *4*, 7–32.

49. Tohoku NLP Group. *Pretrained Japanese BERT Models*; Tohoku NLP Group: Miyagi, Japan, 2022. Available online: <https://huggingface.co/cl-tohoku> (accessed on 31 August 2023).
50. Kawahara Lab at Waseda University. *Pretrained Japanese RoBERTa Models*; Kawahara Lab at Waseda University: Tokyo, Japan, 2021. Available online: <https://huggingface.co/nlp-waseda> (accessed on 31 August 2023).
51. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised cross-lingual representation learning at scale. *arXiv* **2019**, arXiv:1911.02116.
52. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
53. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
54. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
55. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 21 June–18 July 1967; Volume 1, pp. 281–297.
56. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
57. Organisciak, P.; Acar, S.; Dumas, D.; Berthiaume, K. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Think. Ski. Creat.* **2023**, *49*, 101356. [[CrossRef](#)]
58. Reif, E.; Yuan, A.; Wattenberg, M.; Viegas, F.B.; Coenen, A.; Pearce, A.; Kim, B. Visualizing and measuring the geometry of BERT. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2019; Volume 32.
59. McCoy, R.T.; Min, J.; Linzen, T. BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance. In Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Online, 16 November 2020; pp. 217–227. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.